

Running head: HEALTH ENROLLMENT ASSESSMENT REVIEW

The Health Enrollment Assessment Review (HEAR):
Its impact on utilization

Captain William D. Judd
Graduate Management Project

U.S. Army-Baylor University Graduate Program in Health Care
Administration

A paper submitted in partial fulfillment of the
Requirements for HCA 5661 Administrative Residency

14 April 2000

| Report Documentation Page | | |
|--|-------------------------------|--|
| Report Date 00 Aug 2000 | Report Type N/A | Dates Covered (from... to) - |
| Title and Subtitle The Health Enrollment Assessment Review (Hear): Its Impact on Utilization | Contract Number | |
| | Grant Number | |
| | Program Element Number | |
| Author(s) | Project Number | |
| | Task Number | |
| | Work Unit Number | |
| Performing Organization Name(s) and Address(es) Department of Defense Health Service Region 5 Bldg 6, Area B 2776 C Street, Suite 200 Wright Patterson AFB, OH 45433 | | Performing Organization Report Number |
| Sponsoring/Monitoring Agency Name(s) and Address(es) | | Sponsor/Monitor's Acronym(s) |
| | | Sponsor/Monitor's Report Number(s) |
| Distribution/Availability Statement Approved for public release, distribution unlimited | | |
| Supplementary Notes | | |
| Abstract | | |
| Subject Terms | | |
| Report Classification unclassified | | Classification of this page unclassified |
| Classification of Abstract unclassified | | Limitation of Abstract UU |
| Number of Pages 63 | | |

Acknowledgments

I would like to acknowledge the staff at DoD Health Service Region 5 for their assistance in this project. Their contributions helped put this project over some major hurdles. Additionally, I would like to thank Dr. Kenneth Finstuen for his help with the statistical analysis of this project.

Abstract

The purpose of this study was to analyze the Health Enrollment Assessment Review (HEAR) program to determine its impact on utilization. Additionally, the HEAR program is lacking useful metrics to track the success of the system. By conducting comparisons and developing predictor models, this study enables the HEAR program to have meaningful measures and accurately affect changes in the system.

A series of six predictor models were developed to look at the six utilization variables of outpatient visits, emergency room visits, inpatient days, pharmacy prescriptions, laboratory procedures and radiology procedures. This analysis compared two groups of beneficiaries. Those that successfully completed the HEAR (HEAR group) and those that have not completed the HEAR (No HEAR group).

This study randomly selected a sample population of 748 Prime beneficiaries enrolled to the Wright Patterson Medical Center. The HEAR and No HEAR groups each had 374 beneficiaries in the respective groups. The study collected utilization data and demographic characteristics on these beneficiaries for a 12-month period and subsequently analyzed it for differences and development of predictor models.

The comparison of utilization between the HEAR and No HEAR groups yielded no statistically significant differences for five out of six variables. The only significant difference was realized within the emergency room visits category ($p < .05$). Additionally, six predictor models were developed for use in

utilization prediction. Results showed that four of the six models (outpatient visits, laboratory procedures, radiology procedures, and pharmacy prescriptions) produced R^2 that were significant at $p < .001$.

These results have revealed two important findings. First, the HEAR program is not reducing most forms of utilization for its beneficiaries. This allows managers to now develop metrics and set objectives for future changes to the HEAR system. Second, although significant differences do not exist, it is still possible to predict beneficiary utilization based on completion of the HEAR and certain demographic characteristics. This information is useful for HEAR managers and the successful evolution of the HEAR program.

Table of Contents

| | |
|------------------------------------|----|
| List of Tables | 6 |
| List of Figures | 7 |
| Section | |
| 1. Introduction | 8 |
| Conditions | 8 |
| Problem Statement | 13 |
| Literature Review | 14 |
| Purpose | 28 |
| 2. Method and Procedures | 28 |
| Variables | 29 |
| Sample Selection | 35 |
| Data Collection and Analysis | 35 |
| Ethical Considerations | 39 |
| 3. Results | 40 |
| Descriptives and Comparisons | 40 |
| Correlation | 42 |
| Utilization Models | 43 |
| 4. Discussion | 46 |
| HEAR/NoHEAR Comparison | 46 |
| Utilization Models | 48 |
| 5. Conclusions and Recommendations | 50 |
| Metric Development | 51 |
| Model Usage | 52 |
| Study Limitations | 53 |
| Future Research | 54 |
| References | 56 |
| Appendices | 62 |
| A. Operational Definitions | |
| B. Models | |

List of Tables

| Table | | Page |
|-------|--------------------------------|------|
| 1 | Demographic Variables | 34 |
| 2 | HEAR and No HEAR Comparison | 42 |
| 3 | Dependent Variable Correlation | 43 |
| 4 | Utilization Models | 44 |
| 5 | Unstandardized Coefficients | 45 |

List of Figures

| Figure | | Page |
|--------|---|------|
| 1 | Health appraisal theoretical development stages | 15 |
| 2 | Health Resource Utilization Level for population and selected sample | 33 |
| 3 | Beneficiary sample size for HEAR and No HEAR groups within demographic categories | 41 |

Introduction

Conditions

The Military Health System has developed an important tool in its effort to improve the health of the beneficiary population. This tool is the Health Enrollment Assessment Review or more commonly known as the HEAR. This study will examine the importance the HEAR has on utilization and develop models that will help future research predict the impacts the HEAR has on utilization.

HEAR Background

The Health Enrollment Assessment Review (HEAR) program is a preventive care system designed to identify high risk behaviors, chronic conditions and needed prevention services of the beneficiary population (Halpern, Murray, Palmer, Reblando, and Rust, 1994). The HEAR provides a medical snapshot based on the beneficiary's input. Its function is to use this input to facilitate a healthier lifestyle for the beneficiary. The program consists of the HEAR form or questionnaire, analysis of beneficiary answers, and reports to the beneficiary and primary care managers.

The HEAR form is a self-administered questionnaire designed to assess the patient's health status and health behaviors. This questionnaire is a multi-page scannable bubblesheet mailed to beneficiaries, 17 years of age and older, upon enrollment in TRICARE Prime. The questionnaire contains 82 questions. The individual sex, lifestyle, and medical history will dictate how many of these questions require answers. The beneficiary mails

the completed form into the TRICARE contractor for his or her region. The contractor scans in the form into the computer system, called Scan Book, and analyzes the beneficiaries' responses. The computer system then produces output based on preset algorithms (R. Baker, personal communication, August 18, 1999, and October 26, 1999).

This output comes in the form of several reports. The first report is called the patient report card. This is a form letter sent to the beneficiary describing possible services needed, counseling regarding risk factors and chronic conditions (Patient Report Card, 1999). The next report is the primary care manager (PCM) report. This provides the PCM with the patient's possible chronic conditions, risk factors and preventive service recommendations (Primary Care Manager Report, 1999). This is the report the PCM should be using to direct and educate the patient to necessary and appropriate services. These first two reports are the most important in the development of a healthier lifestyle for the beneficiary. The PCM and beneficiary can discuss healthy habits and preventive services necessary to bring about behavior changes toward health living.

Two additional reports are also available to the PCMs, the patient audit report and the patient profile report. The patient audit report provides a list of the answers the beneficiary gave for each question (Patient Audit Report, 1999). The patient profile report allows the PCMs to view all their patients' answers from the HEAR survey on one consolidated

report (Patient Profile Report, 1999). The PCMs will then have a composite list of all the answers given by their patients. This collection of reports is supposed to guide the beneficiary and the primary care manager into developing preventive health services, special care needs, or lifestyle changes for the beneficiary.

Wright Patterson Medical Center provides a good example of how the HEAR reports provide the PCMs with a tool to develop necessary preventive services for the beneficiaries. The hospital receives new Primary Care Manager Reports several times a month. The hospital staff divides these new reports into two categories. Those with a Primary Care Level (PCL) of one are in the first category and those with a PCL of two or three are in the second. The PCL is an estimation of the level of primary care complexity the beneficiary may need. A beneficiary with a PCL of one is the least complex and will usually require services from nurse practitioners, physician assistants or general medical officers. A PCL of two means that a moderate level of complexity exists and may require services from family practice and internal medicine physicians. The beneficiary with a PCL of three is the greatest level of complexity. This beneficiary will often require interaction with physician subspecialists (Murray and Halpern, 1996).

At Wright Patterson Medical Center, those beneficiaries with a PCL of one have their reports filed in the medical records and subsequently reviewed with his or her PCM during the next visit. The hospital takes a more proactive role with those

beneficiaries that have a PCL of two or three. The PCM reviews the beneficiary's PCM report along with the medical records. The clinic then schedules the beneficiary a preventive health appointment to discuss with the PCM areas of concern or deficiency. The hospital, by identifying the PCL 2s and 3s, is attempting to interject various types of preventive services and improve the health of the beneficiary (T. Nickle, personal communication, November 12, 1999).

Unfortunately, the HEAR process does not always work as designed. Multiple problems plague the system which prevent the successful completion and execution of the HEAR surveys. The problems can be categorized into two main groups, the beneficiary and the contractor. The problems that occur with the beneficiary group include improperly filling out the form or not filling out the form at all. The contractor problems that occur include an inability to produce results from forms that may have only minor errors. The contractor could manually correct this; but does not. All of these problems result in only 14.28% of the DoD Health Service Region 5 Prime beneficiary population successfully completing the HEAR (R. Baker, personal communication, August 18, 1999, and October 26, 1999).

The HEAR was developed by a collaboration with MEDTAP International, Battelle Memorial Institute, and the Office of Prevention and Health Services Assessment (OPHSA) of the United States Air Force. The development of the HEAR evolved from several military and civilian Health Risk Appraisal (HRA) systems. The HEAR is different from traditional HRAs because

these HRAs fall short in areas of resource utilization classification, level of care, and preventive service requirements (Office for Prevention and Health Services Assessment, 1999). Additionally, the HEAR does not take the same approach as many HRAs by developing probabilities that describe a persons chance of dying or contracting a disease within the next 10 years.

Department of Defense Health Service Region 6 performed the HEAR pilot test from September 1994 through October 1996 (Murray and Halpern, 1996). All Health Service Regions have since implemented this program. This study focuses on the HEAR program in DoD Health Service Region 5. In particular, the Prime beneficiaries at Wright Patterson Medical Center, Wright Patterson Air Force Base, Ohio were used as the study population.

HEAR Purpose

The HEAR program has four primary objectives: (1) Identify beneficiaries requiring prevention services, (2) Assigning beneficiaries a primary care level (PCL) based on the complexity of needed medical care, (3) Classifying beneficiaries into a predicted level of medical resource utilization, and (4) Identifying beneficiaries with high risk behaviors who could benefit from counseling and health promotion activities (Murray and Halpern, 1996). Conversely, there is no goal of the HEAR system to predict future medical expenditures (Halpern et al., 1994). Instead of predicting a beneficiary's future medical utilization, it takes past utilization and chronic conditions to

put beneficiaries into utilization levels. These levels will help the PCMs to determine the most appropriate avenues of care for the beneficiary. Unfortunately, it does not provide the PCM or facility with a means to predict its population's future use of resources.

Problem Statement

There is currently no system in place to determine if the HEAR program is benefiting the beneficiary population or the Military Health System (MHS) as a whole. The current metrics used to measure the HEAR success only revolve around completion rates. The number of successfully completed HEAR forms based on the number of eligible beneficiaries is the only metric used to measure the health of the system (Technical Outcome Metric #3, 1999). This metric falls short of showing how the HEAR program is reaching its ultimate objectives. In order for the HEAR program to be a benefit to the military health system there must be some other measure for success.

The ultimate goal of the HEAR, as well as many preventive services, should be to improve the health of the individual as well as the overall population. However, these programs can achieve other measurable objectives in the areas of health care costs, or health care utilization (Lauzon, 1977; SRA International, Inc., 1999). A measure in one or more of these areas will result in a better indication of the overall success or failure of the HEAR. Current limitation with the military health system's data collection systems makes population health improvement and health care costs nearly impossible to analyze.

Therefore, this study chose the remaining goal of reducing health care utilization as the essential indicator to determine the level of success for the HEAR system.

Literature Review

Health Appraisal Foundations

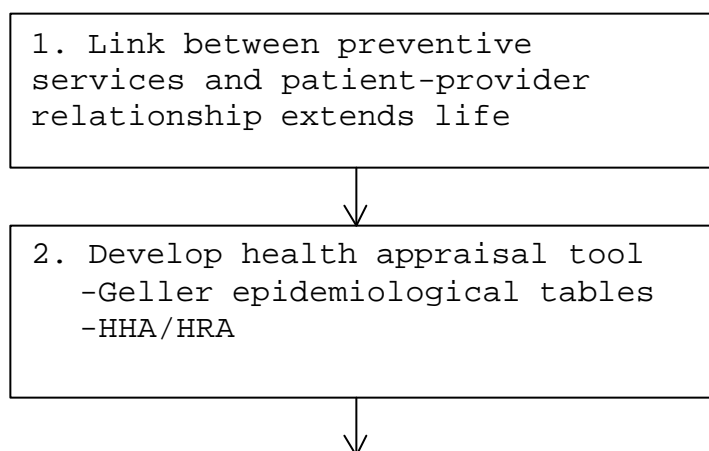
The HEAR's predecessors are two health evaluation systems known as the Health Risk Appraisal (HRA) and the older Health Hazard Appraisal (HHA). The HHA and the HRA are synonymous terms for the same type of health appraisal system. Literature reviews from the 1960s to the 1990s reveal that researchers primarily use the term HHA before the 1980s. A gradual transition occurs in the 1980s with the term HRA primarily used today.

The development of the HHA evolved from the lessons learned during World War II. It became apparent during this time that disease prevention and the importance of the individual practitioner had a dramatic role in increasing life expectancy. The medical profession further built on this concept in the early 1950s and established the link between preventive medicine and the individual patient-doctor relationship (Sadusk and Robbins, 1968).

There are numerous definitions of HHAs/HRAs, but Fielding (1982) provides the best working explanation of the concept. His definition also outlines the purpose of this preventive medicine tool and highlights the limits of the appraisal. Fielding describes the appraisal as a method that determines an individual's chance of dying or acquiring a certain disease type

within a given time frame. This time frame is normally ten years and is based on epidemiological tables established in 1963 by Mr. Harvey Geller of the Cancer Control Program, National Center for Chronic Disease Control of the U.S. Public Health Service (Sadusk and Robbins, 1968). Fielding also went on to explain that the appraisals are limited by the fact that it provides probabilities and not diagnoses. This is important for the individual practitioner and how he or she can best use the tool to help the patient. It does not give the provider a guarantee of disease or illness but only a level of chance that a condition might occur.

Figure 1 displays the staged theoretical evolution of the health appraisal system over the last half-century. These stages greatly overlap from a historical perspective. However, the figure shows a practical prospective of how the appraisals have developed into its current form as well as possible future evolution. This figure also helps to visualize how many of the studies noted in this paper fit into the theoretical constructs of the health appraisal.



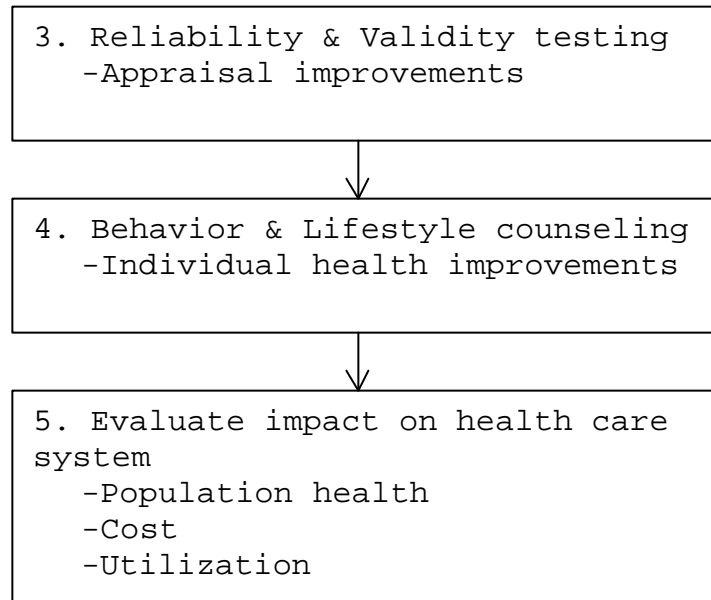


Figure 1. Health appraisal theoretical development stages.

LaDou, Sherwood, and Hughes (1975, 1979) produced two studies that provide good examples of how the appraisal method works and the possible benefits of successful application. These researchers administered a health hazard appraisal to a large group. The appraisal in the LaDou et al. example consisted of a six-page booklet with questions for the patient to complete. The questions included areas such as lifestyle, known diseases, family history, habits, emotional status, and racial background. The answers to these questions were then put with results from a physical examination. The physical examination included a chest x-ray, electrocardiographic studies, submaximal stress electrocardiographic testing, laboratory profiles, and a hands-on physical examination. Researchers then used these results to determine the patient's chances of dying within the next ten years. As a comparison,

the Geller tables describe an average 45 year old white male as having a 9,200 in 100,000 chance of dying in the next ten years (LaDou et al., 1975).

LaDou et al. (1975, 1979) then took the results of the appraisal and developed different age categories for the patient. The researchers developed three ages for each individual completing the appraisal. The first age is the subject's present age. This of course is the individual's age based on the date of birth. The next age is the risk age. The risk age is more than likely higher or lower than the present age. The appraisal tool calculates the risk age based on disease potential, current diseases, lifestyle, and demographic considerations. A healthy 45-year-old male may have a risk age of only 37 years based on a healthy lifestyle and low probability of disease. The final age category is the potential age. This is the possible age the individual can lower his or her risk age to with changes in some controllable factors. For example, a 45-year-old male with a risk age of 50 years stops smoking. This may result in him having a potential age of 46 years.

This is another area where the HEAR differs from its HRA predecessors. The HEAR does not attempt to develop a risk age nor provide the patient with a potential age based on lifestyle changes. Instead, lifestyle changes are encouraged through written and PCM feedback.

LaDou et al. (1975, 1979) also took the next logical step in the health appraisal system and initiated behavior

modification procedures. When providing the patient with the potential age, the researchers also educated the patient on possible lifestyle changes and specific treatments to live a healthier life. The researchers also encouraged the patient to consult his or her own physician for continued care. LaDou et al. showed that this system resulted in an overall reduction in risk age by 1.4 years over a one year period and a reduction in risk age by 2.38 years over four years.

Concerns over HHA/HRA Methodology

These LaDou et al. (1975, 1979) studies highlighted the need for comprehensive counseling that will result in positive behavior changes. Goetz and McTyre (1981) continued this concept by examining the methodology of the appraisal system and its place in the reduction of risk and changes in behavior. These researchers highlighted that the HHA/HRA tools are not solely instruments designed to change behavior. The HHA/HRA should also combine with scientifically proven risk reduction programs and preventive medicine measures. Wagner, Berry, Schoenbach, and Graham (1982) also echo this methodological consideration. Wagner et al. goes on to discuss other concerns about the appraisal systems usage. They describe the popularity of the systems and its contributions to health promotion and disease prevention. However, concerns linger over many of the unproven assumptions that still surround many of the original disease and prevention conclusions. Sadusk and Robbins (1968) highlighted this problem in some of their original work on the appraisal system. They discussed the lack of quantitative

benefit to many of the prevention programs that were universally accepted. More importantly, they noted that concrete data was not available during the development of the original appraisals. In place of concrete data, a committee of experts used its best clinical judgment to produce a needed factor. These factors still existed in many of the HHAs/HRAs of the early 1980s (Goetz and McTyre, 1981).

This raised much concern in the early 1980s about reliability of the HHA/HRA systems. Sacks, Krushat, and Newman (1980) were some of the first researchers to analyze the reliability of the HHA/HRA. Their results seriously questioned the reliability of the appraisal questionnaire. They showed that inconsistent patient responses to the appraisal questions dramatically reduced the effects the tool could have on reducing overall risk.

Problems such as this prompted Goetz and McTyre (1981) to design a methodological framework in which researchers can structure and evaluate the appraisals. The general model developed by these researchers included areas of individual data collection, risk data assumptions, risk estimation algorithms, feedback to the individual, and the evaluation of the appraisal system as a whole. Two of these important areas worth further note are the areas of feedback and evaluation of the overall reduction of risk. These areas are common themes noted from the beginning with Sadusk and Robbins (1968) and are the next logical steps in the evolution of the appraisal systems. The study by Wagner et al. (1982) also foresaw this need for further

research in the measurement of health risk appraisal's ability to produce health-related behavior change. However, this is the important final step that is lacking in the HEAR. The HEAR follows the Goetz and McTyre framework but does not incorporate an adequate evaluation of the whole appraisal system.

The model developed by Goetz and McTyre (1981) seemed to win approval from the industry. In a letter to the editor, McDowell (1982) felt that the model developed by Goetz and McTyre provided a valuable description of the appraisal system. Additionally McDowell went a step further, calling for more analysis of the predictive validity of the HRA system.

Researchers began to address these concerns over reliability, validity, and effectiveness in the late 1980s and early 1990s. Foxman and Edington (1987) conducted a study to evaluate the accuracy of the HRA developed by the Center for Disease Control (CDC). In this study, the researchers compared the observed and predicted mortality rates for individuals over a 20-year period. The results showed that the CDC HRA performed rather well, even though the initial predictions are only for a ten-year period. The predicted rates held up under a more demanding 20-year validation. This was an important step to show that the appraisal tool performed to its original specifications.

Another test of the appraisal system came from Smith, McKinlay, and McKinlay (1989) in the form of reliability analysis. This group tested the reliability of four widely used health risk appraisals. The researchers used a test-retest

technique to compare self-reported baseline responses and risk scores. The researchers administered the test-retest on two occasions seven to 12 weeks apart. The results indicated that the responses were generally consistent and reliability generally high. However, some scores still varied greatly among the different appraisal tools. This variability caused concern for the researchers because the tools did not fully distinguish between changes in risk behavior and random reporting errors. Organizations using one of these appraisal tools as part of a prevention program might also have concerns. These results showed that intervention efforts of prevention programs would be difficult to detect and separate from random errors.

These same researchers, Smith, McKinlay, and McKinlay (1991), also performed validity testing of the same four health risk appraisals for coronary heart disease. These four appraisal tools did show modest correlation in the area of coronary heart disease mortality risk as compared to the epidemiological model. This result is more promising and shows how the HRA tools are producing results that are more accurate and evolving into a more effective appraisal tool.

Stein, Lederman, and Shea (1993) again tested reliability. This study is different from some previous ones because it analyzed a tool designed to improve the predictive ability of the health risk appraisal system. The Behavioral Risk Factor Surveillance System (BRFSS) is the tool developed by the Center for Disease Control and Prevention which estimates the prevalence of risk factors of major causes of death in the

United States. The success of this tool will allow for health risk appraisals to more accurately predict possible disease or health related problems.

The Stein et al. (1993) study also used the test-retest technique to examine this new tool. A questionnaire was administered on two occasions, 21 to 44 days apart. Results showed that reliability was generally high across behavior risk factors. This study determined that the BRFSS would be a reliable tool for determining the prevalence of major risk factors. Therefore, these results show that the BRFSS will be valuable for improving an HRAs prediction and the subsequent behavior counseling process.

Gomel, Oldenburg, Simpson, and Owen (1993) provide the next step in the implementation and evaluation of the appraisal system. This group analyzed an HRA and its ability to make behavior changes. These changes in behavior are the essential link to improving the health of individuals and populations (Goetz and McTyre, 1981; Sadusk and Robbins, 1968; SRA International, Inc., 1999). Gomel et al. demonstrated that the HRA in conjunction with behavioral counseling produced significant behavior changes over an HRA with no counseling. This is an important conclusion for the appraisal system because it showed what the LaDou et al. (1975, 1979) studies did not. It showed that individuals can change behavior and a healthier lifestyle accomplished with a fully developed HRA and follow-up counseling.

These recent studies (Foxman and Edington, 1987; Gomel et al., 1993; Smith et al., 1989, 1991; Stein et al., 1993) highlight the most recent phase of evolution in the HHA/HRA cycle. Reliability and validity testing continue to discern the amount of credibility and value these appraisal tools have. Additionally, the appraisals are being tied closer to prevention programs and behavior modification counseling as the concept was first introduced in the LaDou et al. (1975, 1979) studies. As Goetz and McTyre (1981) indicated in their methodological analysis of the HRA, the feedback to the patient is a critical piece to correcting unhealthy lifestyles or lack of preventive care. This feedback then requires follow up to determine if the patient has truly changed lifestyles or sought preventive services. Gomel et al. provided this important link in their findings about behavior counseling and how it could change lifestyles and promote healthy living. However, these studies still do not address the final link in the appraisal system. This last link is the determination of the effectiveness of the HRA system as a whole. These studies have shown the development of the appraisal system to benefit the individual patient but do not show the same relationship to the overall health care system.

Utilization Factors

As discussed earlier, there are several ways to determine the overall effectiveness of the HRA. Researchers can look at population health, health care cost, or utilization (Lauzon, 1977; SRA International, Inc., 1999). A necessary component for

an appropriate study of population health is the requirement of 10 years of data (Foxman and Edington, 1987). Since the basis of HRAs are predictions over a 10 year period, at least 10 years of data will be necessary to study the HEAR. Since the HEAR is only a few years old, another study will have to address the issue of population health after the system has been in place for at least 10 years. Cost is also not a focus in this study because of the limitations noted in the data collection systems. Therefore, this study has chosen the utilization construct as its basis for analysis of the military's HRA system. Lauzon was first to make this connection between health appraisals and health care utilization. He included utilization in his study of HHAs and lifestyle change. However, his study only took into account doctor visits and only for a 12-week period. The results of his study did not show any statistical differences in the utilization of his three study groups. Lauzon noted that a long-term study, of a year or longer, would be more definitive in demonstrating change.

Following Lauzon (1977), there were several other studies that demonstrated that utilization is an appropriate measure of the effectiveness of HRAs. A study by Mason, Bedwell, Zwaag, and Runyan (1980) makes the best connection between behavior and utilization. These researchers determined that personal habits and lifestyle accounted for 78 percent of preventable hospital admissions. Factors such as alcohol abuse and smoking caused hospital admissions that could have otherwise been prevented. These factors are the same type of unhealthy behaviors that the

HRAs attempt to identify and correct through behavior counseling. Additionally, a study by Hulka and Wheat (1985) has also highlighted this link between behavior and subsequent utilization and supports the Mason et al. conclusions.

An earlier study by Kisch and Kovner (1969) has also shown that a relationship exists between a person's current health status and how much health care he or she will use. This is precisely what the HRA is trying to do on the individual level. The appraisal should develop a profile of an individual that outlines the patient's lifestyle habits, current health status and predictions of future disease. This will then allow for the development of behavior counseling which will result in lifestyle changes and a subsequent reduction in utilization.

A recent article by Frye (1998) highlights a further need to analyze subsequent reductions in utilization and cost savings from prevention services. He states that it is unknown exactly how much preventive care will reduce hospitalization and health care cost. He does note the correlation between preventive care and reduced probability of hospitalization. The Frye article shows the need to find mechanisms to measure the effectiveness of HRAs. The HEAR is no exception to this need. The focus on utilization impacts will be the first step in this final measurement process and determination of the level of success for the whole appraisal system.

One of the caveats in the HEAR program, concerning utilization, is that it does not attempt to predict future utilization. Instead, it merely identifies individuals who are

likely to be high utilizers of medical services (Halpern et al., 1994). This process of identifying high utilizers stops far short of accurately predicting future utilization as well as measuring the HEAR's overall impact on utilization. This shows the need to measure the impact on utilization and develop useful predictor models based on the HEAR appraisal system.

This void in measuring utilization levels after the tool is in place is not only present in the HEAR program but in much of the literature as well. It appears that analysis of the HRA and HEAR tools do not take the next logical step of implementing operational controls. These operational controls will link the HEAR tool with the strategic goals of the organization. The controls will inform the personnel responsible for monitoring the HEAR if the program is meeting its objectives (Ginter, Swayne, and Duncan, 1998). Therefore, the utilization analysis conducted by this study, will serve as a beginning for the development of operational controls for the HEAR program.

There is a great deal of variation in the literature concerning operationally defining utilization. Meyer, Prochazka, Hannaford, and Fryer (1996) define utilization as total clinic visits, emergency department visits, hospital admissions and length of stay, number of laboratory tests, imaging procedures, and medications dispensed. Wong, Hollenberg, and Charlson (1999) took another approach by defining utilization as the number and costs for laboratory tests, radiology test, specialty consultations, and other miscellaneous services such as vaccinations, tuberculosis

screening, and outpatient procedures. Freeborn, Mullooly, Pope, and McFarland (1990) simplified the definition of utilization as the number of office visits and hospital admissions. Whereas these same four authors in an article five years earlier (McFarland, Freeborn, Mullooly, and Pope, 1985) defined utilization as doctor office visits. These visits consisted of all nonpregnancy services provided on an ambulatory basis whether in the medical office, in the emergency room, at home, or by telephone or letter. Perhaps the best source of utilization definitions comes from Murray and Halpern (1996) in their final report of phase two for the HEAR project. There are 17 factors the HEAR considers in its attempts to classify an individual's utilization. Of these 17 factors, only four apply to utilization as this project is concerned. The remaining 13 factors are only useful in utilization classification and not in the comparison performed in this study. The four factors of relevance are prescription medication, outpatient visits, inpatient visits, and emergency room visits.

All these definitions of utilization lead to the conclusion that there is not one consistent or all encompassing definition. Utilization appears to be a study specific definition varying from project to project. Researchers are left to their experience and better judgment to choose which factors will make up the utilization construct. Presented later in this study is the operational definition of utilization.

Purpose

The purpose of this project is to determine if the implementation of the HEAR has significantly reduced the health care utilization of the study population. The project will look at beneficiary utilization variables in terms of prescription medications, outpatient visits, inpatient days, emergency room visits, laboratory procedures, and radiology procedures. This project will also develop predictive utilization models based on HEAR completion and demographic categorization. The models will be useful in helping researchers analyze the affects of the HEAR on utilization and tailor changes to the system.

The hypothesis of this project is that the beneficiaries that have had the HEAR form successfully make it all the way through the system will have a significant reduction in utilization compared with those beneficiaries that have not completed the HEAR. This can also be stated as a functional relationship, $Y(\text{utilization}) = f(\text{HEAR completion})$. This study will show if the level of utilization is a function of an individual successfully completing the HEAR survey.

Methods and Procedures

A comparison of one group with another is a simple characterization of the purpose of this project. That comparison should discover that one group (those that completed the HEAR) have less medical care utilization than the other group (those that did not complete the HEAR). A study that performed similar group comparisons is the McGann and Bowman

(1990) study. Although this study did not analyze utilization, it did provide a methodological framework in which to perform comparisons.

Even more important than the McGann and Bowman (1990) study is the work Lauzon (1977) did on the HHA's ability to improve individual health. Lauzon conducted a controlled trial comparing lifestyle changes resulting from an HHA. His study compared three groups of patients. The first group was the control group. This group of patients did not receive any HHA. The second group received only the HHA but no lifestyle counseling. The third group of patients received both the HHA and lifestyle counseling. Lauzon statistically compared these three groups to determine behavior changes occurring from interaction with the HHA.

This study follows a similar pattern to the Lauzon (1977) study with a few notable differences. First, only two groups are compared in this study as opposed to three for Lauzon. Second, utilization is the focus of the study instead of the broader category of lifestyle change. Third, this study compares groups over a 12-month period instead of only 12 weeks with Lauzon.

Variables

An explanation of variables and further operational definitions are necessary before proceeding. Appendix A provides a list of all the operational definitions in this study. This report has already defined the construct of utilization as emergency room visits, inpatient days, outpatient

visits, medications prescribed, laboratory procedures, and radiology procedures. Each of these six factors will be a separate dependent variable in the ensuing models. The following definitions of the utilization variables will not include visits due to pregnancy as described in McFarland et al. (1985), preventive medicine services or telephone consultations.

Emergency room visits are defined as the number of times the beneficiary has presented to the emergency room. Inpatient days are defined as the number of days the beneficiary has been admitted to a hospital. Outpatient visits are defined as all nonemergency room visits to a primary care or specialty provider. Medications prescribed are defined as the number of prescriptions written for the beneficiary. Laboratory procedures are defined as the number of laboratory procedures ordered for the beneficiary. Finally, radiology procedures are defined as the number of imaging or radiology procedures ordered for the beneficiary. Appendix B provides the models for each of these six dependent utilization variables. These models will help to clarify the functional relationship tested in this study.

As noted in McFarland et al. (1985), they did not count pregnancy related visits because of its ability to dramatically increase utilization numbers and skew results. Therefore, this study not only deleted pregnancy visits but also removed the beneficiary from the sample if she had any pregnancy-related visits during the 12-month study period. This was necessary because it is too difficult to determine which ancillary care

(pharmacy, lab and radiology) was due to normal medical visits or pregnancy related visits.

Additionally, this study separated all visits coded as preventive. These visits were set aside for separate analysis. It was necessary for the study to do this because preventive visits may also skew the data results. As noted earlier many individuals that complete the HEAR will be given preventive appointments with their PCM. If these visits were not filtered out, then it would be unfairly counted against the HEAR group and analysis results would be less meaningful. All preventive coded visits subsequently removed from both the HEAR and No HEAR groups were further analyzed to determine which group received more prevention visits.

Telephone consults were the final group of appointments removed from the data set. This study only considered visits that were face-to-face with a provider.

The next set of variables that needs further defining is the group membership of the sample. All beneficiaries in the sample fall into one of two groups. The group membership variable is a dichotomous variable defined as HEAR or No HEAR. HEAR is further defined as those beneficiaries that have successfully completed the HEAR form. The HEAR form then made it to the contractor. The contractor successfully scanned it in to the system and results sent to the beneficiary and the primary care manager. The No HEAR group is defined as that group which for whatever reason have not successfully completed the HEAR. The HEAR form may not have been completed by the

beneficiary or it may be lost in the contractors system. The reason is not of importance to this study. Only that the form has not been successfully completed.

It must also be noted that the HEAR membership group should be reflective of the population in terms of resource utilization groups. The HEAR system categorizes beneficiaries into three resources utilization levels: level 1, level 2, or level 3. Level 1 being the lowest utilization group and level 3 being the highest utilization group. A previous study has shown that a sample of 136,422 beneficiaries yielded a breakdown as follows: level 1 - 92%, level 2 - 4%, and level 3 - 4% (Meyer, 1998). Therefore, the sample taken from the HEAR population should be randomized but reflective of the three levels of resource utilization percentages. This stratification is necessary to ensure that a random sample of the HEAR group does not yield a dramatically increased proportion of beneficiaries in any one level. This is an important note because people who have few medical problems are more likely to complete HRA type forms than those with unknown or chronic conditions. Figure 2 shows the actual percentages of the HEAR sample group for this study. The resource utilization levels are consistent with that found in previous studies. Conversely, there is no means to stratify the No HEAR sample into these three levels. The No HEAR group would have to complete the HEAR form in order to determine which utilization level they would fall. That would then defeat the purpose of this study. Although this may appear to be a weakness, it is not. A true random sample of the No HEAR

population should result in the correct percentage representation of the utilization levels.

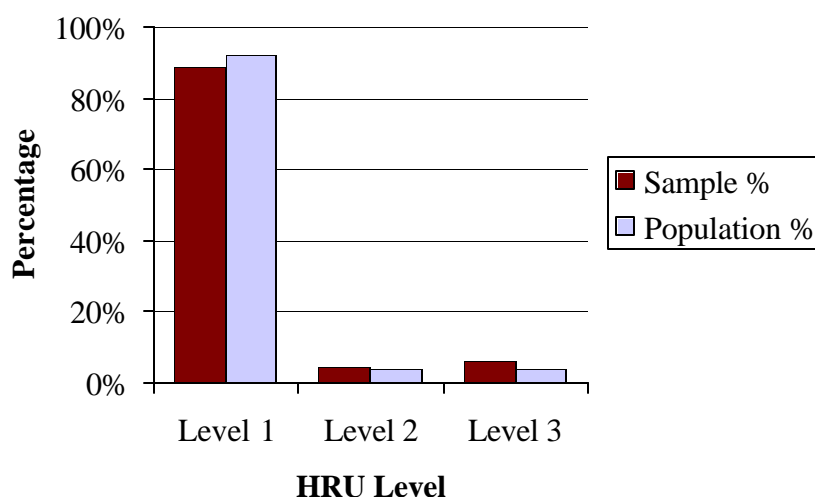


Figure 2. Health Resource Utilization Level for the population and selected sample.

The final set of variables that requires clarification is the demographic breakdown of the sample. Demographics consist of five categories of age, gender, marital status, ethnicity, and beneficiary status. These five categories are further broken out into dichotomous variables. A "1" for a variable indicates the presence of that demographic characteristic. Table 1 indicates the variables found in each demographic category. All the variables, except age, come from the demographic section on the HEAR questionnaire. Age on the HEAR questionnaire is categorized by year and not by class. This study will use age as three separate variables to produce results that are more meaningful. Additionally, there is no age group for 65 and older because this group is not eligible for TRICARE Prime.

Table 1

Demographic Variables

| Category | Variable |
|--------------------|--|
| Age | 17-35 36-45 46-64 |
| Gender | Male Female |
| Marital Status | Married Other than Married |
| Ethnicity | White Black Other |
| Beneficiary Status | Active Duty Service Member(ADSM) Retired Service Member(RSM) Family Member of ADSM Family Member of RSM |

The categories of Marital Status and Ethnicity have consolidated variables called Other. This study created the Other variables in order to consolidate smaller marital status and ethnic groups and include them in the study. This produced variables of sufficient size to perform statistical analysis on. The Other than Married variable consists of beneficiaries from the Divorced, Never Married, Separated and Widowed groups. The Other variable under the Ethnicity category consists of beneficiaries from the American Indian or Alaskan Native, Asian/Oriental, Pacific Islander, Hispanic and Other groups.

Sample Selection

The sample for this study came from DoD Health Service Region 5 at the Wright Patterson Medical Center, Wright Patterson Air Force Base, Ohio. The study limited the sample to those individuals who are TRICARE Prime enrolled to the Medical Center. Many studies (Lauzon, 1977; Smith et al., 1989, 1991) have noted the need for a long-term study of HRAs. This study chose a 12-month study period, to collect utilization data, of September 1998 through August 1999. Therefore, the beneficiaries had to be enrolled in Prime for a minimum of the 12 months of the study period. Additionally, for a beneficiary to be eligible to be in the HEAR sample, he or she must have successfully completed the HEAR survey between May 1998 and August 1998. An example of this is a beneficiary that enrolls in Prime in June 1998 must have successfully completed the HEAR survey by August 1998 and remain in Prime through August 1999 to be eligible for the sample. The study then collected utilization data on this beneficiary for the 12-month study period. The No HEAR group only needed to enroll in Prime to the Medical Center during the May - August 1998 period and remained in Prime for the 12-month study period to be eligible.

Data Collection and Analysis

The sample was identified through two information management systems, the HEAR Data System and the Composite Health Care System (CHCS). These systems identified successful completion of the HEAR, dates of Prime enrollment, and demographic characteristics.

An initial analysis of beneficiaries resulted in several thousand individuals being eligible for the study. A random selection then generated a sample size of 790 beneficiaries. This broke down to 397 for the HEAR group and 393 for the No HEAR group. The study then refined this list further by removing all beneficiaries that had a pregnancy-related visit during the 12-month study period. This resulted in the removal of 42 beneficiaries from the sample. The final results for the two groups were 374 for the HEAR and 374 for the No HEAR group.

One additional consideration taken into account during the sample selection was the possibility of changes occurring in demographic categories. During the 12-month period in which the study collected utilization data, it is possible that some beneficiaries had an event that caused a shift in a demographic status. Such common shifts could come in the form of changes in age category because of birthdays, changes in marital status, or changes in beneficiary status such as a transition from active duty to retirement. This study collected demographic characteristics as it existed at the beginning of the 12-month data collection period. These characteristics were then locked-in. No attempt was made to compare demographic characteristics at the beginning and end of the 12-month data collection period. Therefore, this study made no adjustments in demographics based on changes during the data collection period.

The Corporate Executive Information System (CEIS) provided utilization data for the beneficiaries in the sample. CEIS is a computer program capable of allowing the user to access an

enormous warehouse of health care data. The program draws data from many source computer systems located at the hospital level. The two main source systems that provide CEIS with utilization data are the Composite Health Care System (CHCS) and the Ambulatory Data System (ADS). CHCS and ADS provide various means of collecting inpatient and outpatient data as well as prescriptions and ancillary procedures. The systems feed this information into the data warehouse where CEIS can access it. This provides users with an effective means to obtain data from multiple source systems.

The CEIS data pulled for this study included utilization of both the military treatment facility and the community providers. The data were pulled from the computer systems available at the DoD Health Service Region 5 Lead Agent Office, Wright Patterson Air Force Base, Ohio.

The study statistically analyzed the data to determine the significance of relationships and models. Descriptive statistics were isolated for the HEAR and No HEAR groups across the six utilization variables. Additionally, the study performed correlations between the six utilization variables (Outpatient visits, Emergency room visits, Inpatient days, Prescriptions, Laboratory and Radiology procedures) to determine the interrelationship between the models. The study also generated and analyzed six models to determine utilization predictability from group membership and demographic characteristics. The study used an F-test to determine the significance of comparisons and models.

Two other relevant studies, Wong, Hollenberg, and Charlson (1999) and Meyer, Prochazka, Hannaford, and Fryer (1996), are examples of the form of analysis used in this study. Although neither study directly discusses HRAs or preventive services, both studies do discuss utilization comparisons. These studies provide models that compare utilization between group memberships. They describe the type of statistical test used in their analysis. Additionally, an article by Brooke, Hudak, and Finstuen (1994) provides a framework for model development and analysis applicable to this study.

The alpha probability level for all models of the analysis is $\alpha = .05$. Additionally, the study took into consideration the validity and reliability of the results. Validity was determined using R^2 and measured the internal validity of the six individual models. This coefficient displays the amount of shared variance accounted for in the model. The higher the coefficient the more variance accounted for and the more accurate the model.

A study is not complete without analysis of the consistency of its measurements. The literature reveals several different methods researchers used to measure reliability of HRAs. These methods include Cronbach's Alpha (Brazier et al., 1992), correlations (Smith et al., 1989) and Kappa (Stein et al., 1993). Brazier et al. used Cronbach's Alpha to measure the internal consistency of the SF-36 health survey questionnaire. Smith et al. used correlations to determine the reliability of the test-retest technique for four different HRAs.

Additionally, the Stein et al. study took the approach of using kappa and unweighted kappa to determine the reliability of categorical variables. Since the approach taken by Brazier et al. mirrors the approach of this study, this study has chosen Cronbach's Alpha as its measure of reliability. A random sample of beneficiaries who completed the HEAR was chosen to determine the level of reliability.

Ethical Considerations

This study conforms to all ethical and legal guidelines established within the industry and the institution. This study sought the approval of the Wright Patterson Medical Center Institutional Review Board (IRB) before the data collection phase. The IRB approved the project and the collection of necessary data.

The methodology of this study ensured handling of patient information in a very sensitive manner. The study did not use identification data in this or other reports. Patients' names and social security numbers were only used to collect utilization data and ensure that duplicate records were not present in the data set. Once the study eliminated all duplicate records and collected utilization data, it deleted the names and social security numbers from the data set. After the deletion of these fields, there were no patient identifiable systems in the data set. Additionally, patient information was aggregated within the described groups. The study did not display individual patient data in any report. Results are only displayed as a population, sample or demographic group.

Results

Comparisons of the HEAR and No HEAR groups resulted in statistically significant differences in only one utilization variable (emergency room visits, $p < .05$). The remaining five variables showed no differences. However, four predictor models (outpatient visits, laboratory procedures, radiology procedures and pharmacy prescriptions) produced significant R^2 ($p < .001$).

Descriptives and Comparisons

When selecting the sample for this study, it became apparent that demographic representation be similar for the HEAR and No HEAR groups in order to make valid comparisons (Cooper and Schindler, 1998). This is difficult to accomplish because of the random sampling process utilized by this study. The researcher did not have control over demographic representation. Therefore, the similarities produced by the sampling were the result of chance alone.

Figure 3 shows the breakdown of the fourteen demographic groups for the HEAR and No HEAR sample. Similar representation did exist within most of the demographic groups. It appears that random sampling did yield study groups that possess similar demographic characteristics. One exception to this is the Active Duty beneficiary status group. A random selection process yielded 74 active duty for the HEAR group while only 20 active duty for the No HEAR group. Since this was the result of a random selection, the study did not attempt to adjust the sample size of active duty members. A sample of 20 active duty

is adequate for this study to make comparisons and develop models.

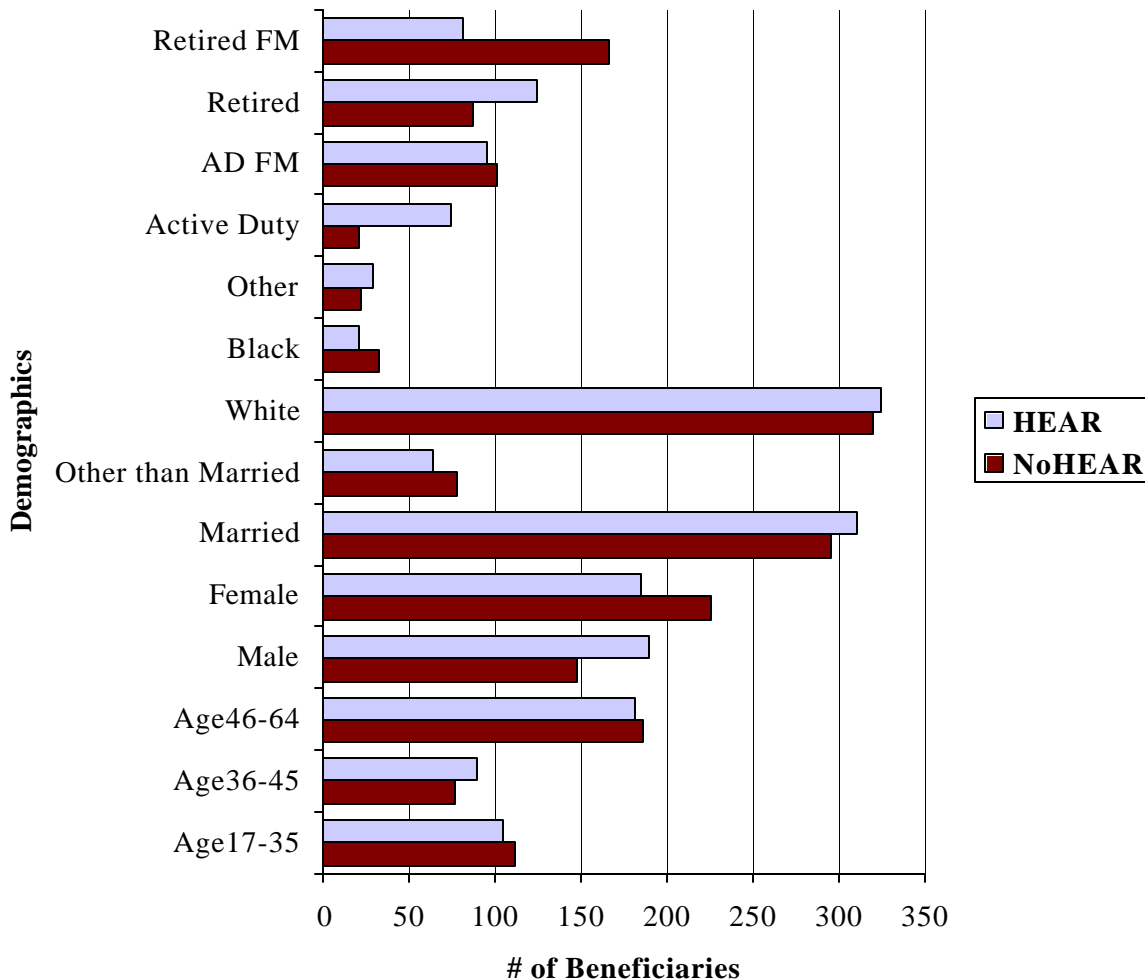


Figure 3. The number of beneficiaries by group membership within each of the fourteen demographic variables.

Unexpected findings arose with the comparison of utilization values for the HEAR and No HEAR groups. Unfortunately, these results are not consistent with the hypothesis of this study. Table 2 highlights the findings of this comparison. Each group had a sample size of 374 beneficiaries. The only utilization variable that produced

expected results was the emergency room visits ($p < .05$). This variable showed that the HEAR group had significantly fewer emergency room visits than did the No HEAR group. All other variables showed that the HEAR group had on average higher utilization or no significant differences than the No HEAR group.

Table 2

HEAR and No HEAR Comparison

| Utilization | HEAR | | No HEAR | | F | p |
|-------------------|------|-------|---------|------|------|-------|
| | Mean | S.D. | Mean | S.D. | | |
| Outpatient Visits | 6.70 | 9.88 | 6.11 | 7.55 | .837 | .361 |
| ER Visits | .39 | .90 | .54 | .99 | 5.19 | .023* |
| Lab Procedures | 6.23 | 9.74 | 6.21 | 9.85 | .001 | .982 |
| Rad Procedures | 1.20 | 2.23 | 1.24 | 2.28 | .051 | .821 |
| Pharmacy | 7.70 | 13.11 | 7.55 | 9.10 | .033 | .856 |
| Inpatient Days | .33 | 1.86 | .24 | 1.52 | .569 | .451 |
| Preventive Visits | .84 | 1.28 | .68 | 1.48 | 2.68 | .102 |

Note. *Significant at $p < .05$.

Correlation

It is necessary to show the correlation between the six dependent variables of the models in order to have a measure of external validity for these predictor models (Cooper and Schindler, 1998). This study chose Pearson's Correlation to measure this level of external validity. Table 3 shows the

correlation between the six utilization variables that have a significant interrelationship. This ranges from a high Pearson's Correlation of .721 between Pharmacy and Lab to a low correlation of .168 between Outpatient visits and Inpatient days. The significance of these relations is expected and consistent with the findings in Roos N., Roos L., Mossey and Havens (1988). Their analysis of various forms of utilization also revealed correlations that ranged from a high of .93 to a low of .17.

Table 3

Dependent Variable Correlation

| | | Outpt | ER | Lab | Rad | Phar |
|-----------|----------------|-------|------|------|------|------|
| ER | Pearson Corr | .262 | | | | |
| | Sig (2-tailed) | .000 | | | | |
| Lab | Pearson Corr | .530 | .354 | | | |
| | Sig (2-tailed) | .000 | .000 | | | |
| Rad | Pearson Corr | .552 | .358 | .637 | | |
| | Sig (2-tailed) | .000 | .000 | .000 | | |
| Phar | Pearson Corr | .557 | .379 | .721 | .537 | |
| | Sig (2-tailed) | .000 | .000 | .000 | .000 | |
| Inpt Days | Pearson Corr | .168 | .195 | .487 | .463 | .451 |
| | Sig (2-tailed) | .000 | .000 | .000 | .000 | .000 |

Note. N=748

Utilization Models

As described earlier, this study developed six dependent predictor models. The purpose of these models was to show that given group membership and certain demographic characteristics,

various types of utilization could be predicted. The results of these models yielded expected findings. Four out of the six models demonstrated significant R^2 with 10 and 737 degrees of freedom. Table 4 shows the results of the six models. The ancillary services of laboratory, pharmacy and radiology have the highest R^2 and predictive validity. Table 5 further defines the models by listing the unstandardized coefficients of only the four significant models. The table shows the independent predictor variables with the necessary coefficients for model implementation.

Table 4

Utilization Models

| Models | R^2 | df1 | df2 | F | p |
|-----------|-------|-----|-----|-------|------|
| Outpt | .064 | 10 | 737 | 5.024 | .000 |
| ER | .019 | 10 | 737 | 1.412 | .170 |
| Lab | .117 | 10 | 737 | 9.735 | .000 |
| Rad | .073 | 10 | 737 | 5.844 | .000 |
| Phar | .092 | 10 | 737 | 7.502 | .000 |
| Inpt Days | .019 | 10 | 737 | 1.405 | .173 |

Table 5

Unstandardized Coefficients

| Predictor Variable | Outpatient | Lab | Pharmacy | Radiology |
|--------------------|------------|--------|----------|-----------|
| Constant | 7.394 | 10.096 | 9.957 | 1.433 |
| GrpMembership | -.555 | -.053 | -.490 | .029 |
| Age1 | -3.700 | -6.448 | -5.641 | -1.025 |
| Age2 | -2.832 | -5.762 | -5.414 | -.818 |
| Gender | 3.758 | 2.584 | 3.474 | .284 |
| MarStatus | .050 | .090 | -.500 | -.237 |
| Ethnicity1 | .418 | -.824 | .085 | .406 |
| Ethnicity2 | -1.530 | -2.152 | -1.610 | .027 |
| BeneStatus1 | -.748 | -1.002 | -1.201 | -.320 |
| BeneStatus2 | -1.622 | -1.813 | -1.973 | -.367 |
| BeneStatus3 | -2.516 | -2.110 | -1.138 | -.239 |

A random sample of 20 HEAR beneficiaries was selected to perform reliability analysis. The whole HEAR sample was not selected because this is only a check of a previously established instrument. The Cronbach's Alpha for these 20 beneficiaries was .9336 and shows a high level of survey consistency. This is also consistent with a previous study by Brazier et al. (1992); where they achieved a Cronbach's Alpha of .85.

Discussion

The results of this study are a mix of expected and unexpected findings. The impact the HEAR process has on utilization appears to be minimal. Conversely, four models do show significant results in its ability to predict utilization. Regardless of the level of impact, a useful metric has been established for the HEAR system. The findings of this study outline the development of this metric and guide HEAR managers to establishing a means to determine the health of the HEAR program.

HEAR/No HEAR Comparison

The results of this study have shown that the beneficiaries completing the HEAR are not showing any less utilization than those that did not. The only exception to this is emergency room utilization. This shows that the HEAR is not having much of an impact on most forms of utilization. The small differences in the two groups reveal that the beneficiaries completing the HEAR and any subsequent preventive visits are not changing the need to utilize the health care system. One out of six utilization comparisons is not strong enough to warrant acceptance of this studies hypothesis. The HEAR system is not producing a reduction in most types of utilization for those that successfully complete the HEAR.

The differences this study noted in the emergency room utilization might result from patient awareness of medical care access procedures. Perhaps, the beneficiaries that completed the HEAR may be more aware of the TRICARE system and how to

access the medical system through means other than the emergency room. This is speculative at this point and would require further research to draw conclusions that are more definitive.

The studies by Gomel et al. (1993) and Lauzon (1977) noted an HRA/HHA ability to change behavior over the short-term. However, the results of this study show that this change does not transfer to the long-term impact on utilization. The behavior changes noticed by the researchers are not resulting in similar changes in utilization. This would lead to the conclusion that short-term behavior changes are not leading to long-term impacts in utilization.

This study's results are consistent with the limited research Lauzon (1977) performed on utilization. Lauzon's findings over a 12-week period are consistent with this study's findings over a 12-month period. Lauzon showed that there was no statistical difference among test and control groups in the number of visits made to a physician or to a health unit. His conclusions showed that although no difference existed in a short-term study, perhaps a longer study might yield different results. This study now shows that the impact over the long-term does not show significant differences.

Another noticeable result is the lack of significant differences between the two groups in the area of preventive visits ($p < .102$). At initial glance it would appear that those beneficiaries that completed the HEAR would create more preventive visits than those beneficiaries that did not. This would be especially true for those beneficiaries from Wright

Patterson Medical Center. The structure of its HEAR program to have Primary Care Level (PCL) 2 and 3 patients come in for a preventive visit would naturally create more visits for the HEAR group. Unfortunately, the hospital offers these beneficiaries a preventive visit but not all accept the visits. Recent analysis determined 2887 beneficiaries with PCL 2s and 3s were contacted and offered preventive visits. Only a third of these patients actually accepted a preventive exam (T. Nickle, personal communication, November 12, 1999).

This low acceptance of preventive visits could account for a lack of significant differences in the preventive visit comparison. It explains why the HEAR group has a slightly higher mean (.84) of preventive visits than the No HEAR group (.68). Significant results may arise if a larger proportion of PCL 2 and 3 beneficiaries accepted the hospital's offer for a preventive visit.

Utilization Models

The significant correlation among the six utilization variables is an important component to the development of the six predictor models. This correlation shows the interrelationship of the six utilization models and reveals the level of external validity across different settings. It reveals that the models ability to generalize for the population is sound. It also adds strength to the predictability of the ensuing models and shows that there is common variability across the models. Therefore, a change in one utilization variable will result in changes in the other utilization variables.

This leads to the second purpose of this study, which was to development predictor models for six different types of health care utilization. This study has shown that four types of utilization, outpatient visits, pharmacy prescriptions, laboratory procedures, and radiology procedures, produced significant R^2 ($p < .001$) and therefore can be predicted by knowing group membership and demographic characteristics. Although no differences exist between group membership, membership and demographic variables together contribute enough predictability to establish a valid model. The remaining utilization models of emergency room visits and inpatient days do not have adequate predictors in the equation to produce statistically significant models (Emergency room: $p < .170$, Inpatient days: $p < .173$). There is undoubtedly much more variability in these two variables than the four that produced significant results. This additional variability needs to be identified and added to the emergency room and inpatient models before any significant predictability could be explained.

Additionally, this study did not attempt to produce predictor models for preventive visits. Preventive visits traditionally behave differently compared to other forms of utilization. Hulka and Wheat (1985) noted this in their study of utilization patterns. These researchers stated that preventive visits are not influenced by the same factors as other types of utilization. This fact makes predicting preventive visits very difficult.

The R^2 for the four significant models ranged from .117 to .064. This shows how much variability the models take into account. Although the models take into account only 6.4% to 11.7% of the variability, this is consistent with other studies in the area of health care utilization. Studies produced by Kosloski, Austin, and Borgatta (1987), Wolinsky, Coe, Mosely, and Homan (1985), and Roos et al., (1988) yielded significant R^2 that ranged from .05 to .23. This study did not expect the R^2 to be higher because of the large variability associated with utilization. There are no hard or fast rules with the prediction of utilization variables. The predictors in this study do not exhaust the range of other potential predictors. However, a study by Kosloski et al. does show that a few key variables can have reasonable validity for utilization predication. Since it is not possible to have all factors associated with utilization in this model, the resulting R^2 of the four significant models is at acceptable levels. This is further evidence of the construct validity of the models and adds strength to the accuracy of its utilization predictability.

Conclusions and Recommendations

This study has successfully developed four utilization predictor models based on a beneficiary successfully completing a HEAR survey and other demographic variables. It has established a method for HEAR program managers to implement quality metrics that will show the true health of the HEAR system. Unfortunately, this new metric has shown that the HEAR

is not producing significant changes in most form of utilization.

Since significant differences do not exist, this study recommends further evaluation of the HEAR tool. This evaluation should take a system approach to analyze all processes. An analysis may determine that minor or even major changes to the program are necessary. Changes could take many forms such as additional or different beneficiary education, different means of HEAR feedback to the beneficiaries and PCMs, or HEAR related disease management or case management. Incentives tied to the completion of the HEAR may also result in better compliance and completion of the HEAR survey. The comparisons and predictor models established in this study will help program managers to determine appropriate changes and monitor its impacts on the HEAR. The analysis of this report was not of sufficient scope or depth to warrant the replacement of the HEAR with a different health appraisal tool.

This study recommends that researchers perform similar HEAR studies in other Military Treatment Facilities (MTFs) and other Health Service Regions. The results of these tests will allow researchers to compare and analyze a broad set of tests and better tailor adjustments to the HEAR system. This will allow researchers to determine trends and adjust the HEAR system to meet the objectives it is designed to produce.

Metric Development

The results of this study have established a medium in which HEAR program managers can use to establish a series of

measurable objectives for the effectiveness of the HEAR system. No significant differences between the HEAR and No HEAR group for most types of utilization shows a starting point for metric development. Thresholds can be established based on these results and periodic monitoring can determine if the HEAR system is meeting desired objectives.

The HEAR's current metric of people who have successfully completed the HEAR is a necessary measurement. It is important for managers to know what proportion of the population is completing the HEAR in order to direct change and get the maximum benefit possible from the program. However, as this study has shown, the HEAR system has been in place long enough for more effective metrics to be established. By using utilization as a metric, managers can test the system on an annual basis to determine if the HEAR meets established objectives.

The establishment of measurable objectives for the HEAR program, such as reduced utilization, can be used to effect future change in the program. Without appropriate measurements, program managers will be unable to determine which change if any is necessary for the improvement of the system. Metrics will guide the managers and allow for an improving and evolving appraisal system.

Model Usage

This study has developed four utilization models useful for the prediction of beneficiary utilization. Military treatment facilities can use these models to determine the projected

utilization needs of its beneficiary population. It can further predict utilization based on the number of beneficiaries that complete the HEAR survey and their demographic characteristics. This will assist the facilities in the ability to determine and provide the appropriate utilization avenues for its beneficiary population.

Study Limitations

There are three limitations associated with this study. First, the study does not take into consideration other important measurements such as cost and population health improvements. These types of measurements may produce results that are similar or dramatically different from this study. Although these are good indicators of the success of the HEAR program, the data collection systems of the military health care system do not allow for complete and accurate analysis. An extensive data collection and analysis system would need to be developed and validated before cost and population health improvements can be measured. Also, the HEAR has not been in place long enough to accurately measure changes in long-term population health (Foxman and Edington, 1987). These indicators are still valuable tools to help determine the health of the HEAR system. Aggressive measures should be taken in order to develop systems that can track and analyze the HEAR's impact on cost and population health. This will result in a fully evolved appraisal system capable of evaluation in all essential areas.

The second limitation of this study is the fact that it does not develop methods of application for other health service

regions. The model developed in this region may not be applicable in all other regions. As noted earlier, an analysis of other health service regions would be necessary to develop a more complete and standardized measurement.

The third limitation is associated with data integrity. Accurate and reliable data are always a concern for researchers. The results of this study are only as good as the quality of data in the MHS computer systems. This study has assumed that the data quality of the applicable systems is good. The study did not find any evidence to support the contrary. This assumption goes for both the demographic and utilization data sources. Therefore, this limitation has insignificant impact unless otherwise successfully disputed.

Future Research

This study has begun the process of measuring the effectiveness of the HEAR program and establishes useful metrics. This utilization analysis will give HEAR program managers a better indication of how the appraisal tool is affecting the broader use of the health care system. However, there is still room for much more analysis into the HEAR and its overall impacts.

There are a number of research possibilities that can build on the foundation of this study. The first of which involves even longer-term studies to evaluate the effectiveness of the HEAR. A study such as this would require at least 10 years worth of data and would measure the health of the population. The study could compare two groups base on the completion of the

HEAR. An analysis could be made to determine if the HEAR population is healthier than the No HEAR population after 10 years. This study would require significant research into what variables constitute a healthy population and the appropriate means to measure them.

Researchers may also perform additional studies with improvements in the MHS costing computer systems. This would be valuable information to know if the HEAR program is costing or saving the MHS money. A very detailed cost-benefit analysis would be necessary to perform such a study.

Studies that would further complement this utilization study involve analyzing the severity level of visits and further investigations into the differences of emergency room utilization. A study to look at the variability in utilization could take into account that perhaps the HEAR group would have less resource intense visits than the No HEAR group. The other study could focus on the differences uncovered in emergency room utilization and determine why there would be significant differences in this utilization variable.

This study has broken new ground in the pursuit to measure the impact the HEAR program is having. Further establishment of program goals and objectives can now be formed from these results. Researchers have firmly developed the HEAR program from years of industry work in the appraisal field. This study has taken the next step. It has begun the process to analyze and determine the true impacts of a health appraisal system.

References

Brazier, J. E., Harper, R., Jones, N. M. B., O'Cathain, A., Thomas, K. J., Usherwood, T., & Westlake, L. (1992). Validating the SF-36 health survey questionnaire: new outcome measure for primary care. British Medical Journal 305(6846), 160-164.

Brooke, P. P. Jr., Hudak, J. D., & Finstuen, K. (1994). Product-line evaluation of graduate medical education program costs. Hospital & Health Services Administration, 39(2), 265-278.

Cooper, D. R., & Schindler, P. S. (1998). Business Research Methods (6th ed.). Boston: Irwin/McGraw-Hill.

Fielding, J. E. (1982). Appraising the health of health risk appraisal. American Journal of Public Health, 72(4), 337-340.

Foxman, B., & Edington, D. W. (1987). The accuracy of health risk appraisal in predicting mortality. American Journal of Public Health, 77(8), 971-974.

Freeborn, D. K., Mullooly, J. P., Pope, C. R., & McFarland, B. H. (1990). Smoking and consistently high use of medical care among older HMO members. American Journal of Public Health, 80(5), 603-605.

Frye, R. (1998). Using preventive care to decrease hospital admissions of pediatric patients. Health Care Strategic Management, 16(9), 16-17.

Ginter, P. M., Swayne, L. M., & Duncan, W. J. (1998). Strategic Management of Health Care Organizations (3rd ed.). Malden, Mass: Blackwell Publishers Inc.

Goetz, A. A., & McTyre, R. B. (1981). Health risk appraisal: Some methodologic considerations. Nursing Research, 30(5), 307-313.

Gomel, M., Oldenburg, B., Simpson, J. M., & Owen, N. (1993). Work-site cardiovascular risk reduction: A randomized trial of health risk assessment, education, counseling, and incentives. American Journal of Public Health, 83(9), 1231-1238.

Halpern, M. T., Murray, M. I., Palmer, C. S., Reblando, J. A., & Rust, S. W. (1994, November 22). Final Report, Project HEAR-Health Enrollment Assessment Review, Phase One: Literature Review, Analysis, and Recommendations. Paper submitted to Office of Prevention and Health Services Assessment. [On-line].

Available:

<http://www.phso.brooks.af.mil/hear1/Summary/LitRvw94.doc>

Hulka, B. S., & Wheat, J. R. (1985). Patterns of utilization: The patient perspective. Medical Care, 23(5), 438-460.

Kisch, A. I., & Kovner, J. W. (1969). The relationship between health status and utilization of outpatient health care services. Archives of Environmental Health, 18(5), 820-833.

Kosloski, K., Austin, C., & Borgatta, E. (1987). Determinants of VA utilization. Medical Care, 25(9), 830-846.

LaDou, J., Sherwood, J. N., & Hughes, L. (1975). Health hazard appraisal in patient counseling. The Western Journal of Medicine, 122(2), 177-180.

LaDou, J., Sherwood, J. N., & Hughes, L. (1979). Health hazard appraisal counseling - continuing evaluation. The Western Journal of Medicine, 130(3), 280-285.

Lauzon, R. R. J. (1977). A randomized controlled trial on the ability of health hazard appraisal to stimulate appropriate risk-reduction behavior. Unpublished doctoral dissertation, University of Oregon, Eugene.

Mason, W. B., Bedwell, C. L., Zwaag, R. V., & Runyan, J. W. (1980). Why people are hospitalized: A description of preventable factors leading to admission for medical illness. Medical Care, 18(2), 147-163.

McDowell, I. (1982, November/December). On the validity of health risk appraisal [Letter to the editor]. Nursing Research, 31(6), p. 347.

McFarland, B. H., Freeborn, D. K., Mullooly, J. P., & Pope, C. R. (1985). Utilization patterns among long-term enrollees in a prepaid group practice health maintenance organization. Medical Care, 23(11), 1221-1233.

McGann, K. P., & Bowman, M. A. (1990). A comparison of morbidity and mortality for family physicians' and internists' admissions. The Journal of Family Practice, 31(5), 541-545.

Meyer, J. G. (1998, February). HEAR: A powerful targeting tool for readiness and managed care. Presentation at the TRICARE Conference, Washington D.C. [On-line]. Available: <http://www.phso.brooks.af.mil/heard/slides/dcfef98.ppt>

Meyer, T. J., Prochazka, A. V., Hannaford, M., & Fryer, G. E. (1996). Randomized controlled trial of residents as gatekeepers. Archives of Internal Medicine, 156(21), 2483-2487.

Murray, M. I., & Halpern, M. T. (1996, March 26). Final Report, Project HEAR: Health Enrollment Assessment Review, Phase Two: Development of HEAR Instrument and Result of Field Test.

Paper submitted to Office of Prevention and Health Services Assessment. [On-line]. Available:

<http://www.phso.brooks.af.mil/hear1/Summary/FinRpt96.doc>

Office for Prevention and Health Services Assessment. (1999, September 10). Health Enrollment Assessment Review (HEAR) and the Preventive Health Care System [On-line]. Available:

<http://www.phso.brooks.af.mil/hear1/Slides/4hearphs/4hearphs.ppt>

Patient Audit Report (1999, September 22). [On-line]. Available:

<http://www.phso.brooks.af.mil/hear1/Reports/pataudit.doc>

Patient Profile Report (1999, September 22). [On-line]. Available:

http://www.phso.brooks.af.mil/hear1/Reports/pat_prfl.doc

Patient Report Card (1999, September 22). [On-line]. Available:

http://www.phso.brooks.af.mil/hear1/Reports/PRC1_2.doc

Primary Care Manager Report (1999, September 22). [On-line]. Available:

http://www.phso.brooks.af.mil/hear1/Reports/pcm_rpt.doc

Roos, N. P., Roos, L. L., Mossey, J., & Havens, B. (1988). Using administrative data to predict important health outcomes. Medical Care 26(3), 221-239.

Sacks, J. J., Krushat, W. M., & Newman, J. (1980). Reliability of the health hazard appraisal. American Journal of Public Health, 70(7), 730-732.

Sadusk, J. F., & Robbins, L. C. (1968). Proposal for health-hazard appraisal in comprehensive health care. Journal of the American Medical Association, 203(13), 1108-1112.

Smith, K. W., McKinlay, S. M., & McKinlay, J. B. (1989). The reliability of health risk appraisals: A field trial of four instruments. American Journal of Public Health, 79(12), 1603-1607.

Smith, K. W., McKinlay, S. M., & McKinlay, J. B. (1991). The validity of health risk appraisals for coronary heart disease: Results from a randomized field trial. American Journal of Public Health, 81(4), 466-470.

SRA International, Inc. (1999). Self-reported health information management study. Draft.

Stein, A. D., Lederman, R. I., & Shea, S. (1993). The behavioral risk factor surveillance system questionnaire: Its reliability in a statewide sample. American Journal of Public Health, 83(12), 1768-1772.

Technical Outcome Metric #3 (1999, September 13). [On-line]. Available:
http://pmtsys.usafsg.bolling.af.mil/afmmis/help/TO03_help.html

Wagner, E. H., Beery, W. L., Schoenbach, V. J., & Graham, R. M. (1982). An assessment of health hazard/health risk appraisal. American Journal of Public Health, 72(4), 347-352.

Wolinsky, F. D., Coe, R. M., Mosely, R. R., & Homan, S. M. (1985). Veterans' and nonveterans' use of health services. Medical Care, 23(12), 1358-1371.

Wong, M. D., Hollenberg, J. P., & Charlson, M. E. (1999). A comparison of clinical performance of primary care and traditional internal medicine residents. Medical Care, 37(8), 773-784.

Appendix A

Operational Definitions

| Construct/Variable | Definition |
|------------------------|---|
| Utilization | Emergency Room Visits, Inpatient Days, Outpatient Visits, Pharmacy Prescriptions, Laboratory Procedures, Radiology Procedures per beneficiary |
| Emergency Room Visits | Number of visits to the emergency room |
| Inpatient Days | Number of days admitted to a hospital* |
| Outpatient Visits | Number of outpatient visits** |
| Pharmacy Prescriptions | Number of prescribed medications |
| Laboratory Procedures | Number of laboratory procedures ordered |
| Radiology Procedures | Number of radiology/imaging procedures ordered |
| Group Membership | HEAR or No HEAR |
| HEAR | Beneficiaries that have successfully completed the HEAR form and it has made it all the way through the system and results sent to the Primary Care Manager and beneficiary |
| No HEAR | Beneficiaries that have not successfully completed the HEAR form for whatever reason |

*Does not include admissions due to obstetrical care.

**Does not include visits for obstetrics, preventive services and telephone consults.

Appendix B

Models

-
1. $Y1(\text{ER Visits}) = a_0U + b1\text{GrpMembership} + b2\text{Age1} + b3\text{Age2} + b4\text{Gender} + b5\text{MarStatus} + b6\text{Ethnicity1} + b7\text{Ethnicity2} + b8\text{BeneStatus1} + b9\text{BeneStatus2} + b10\text{BeneStatus3}$
 2. $Y2(\text{Inpt Days}) = a_0U + b1\text{GrpMembership} + b2\text{Age1} + b3\text{Age2} + b4\text{Gender} + b5\text{MarStatus} + b6\text{Ethnicity1} + b7\text{Ethnicity2} + b8\text{BeneStatus1} + b9\text{BeneStatus2} + b10\text{BeneStatus3}$
 3. $Y3(\text{Outpt Visits}) = a_0U + b1\text{GrpMembership} + b2\text{Age1} + b3\text{Age2} + b4\text{Gender} + b5\text{MarStatus} + b6\text{Ethnicity1} + b7\text{Ethnicity2} + b8\text{BeneStatus1} + b9\text{BeneStatus2} + b10\text{BeneStatus3}$
 4. $Y4(\text{Pharmacy}) = a_0U + b1\text{GrpMembership} + b2\text{Age1} + b3\text{Age2} + b4\text{Gender} + b5\text{MarStatus} + b6\text{Ethnicity1} + b7\text{Ethnicity2} + b8\text{BeneStatus1} + b9\text{BeneStatus2} + b10\text{BeneStatus3}$
 5. $Y5(\text{Lab Proc}) = a_0U + b1\text{GrpMembership} + b2\text{Age1} + b3\text{Age2} + b4\text{Gender} + b5\text{MarStatus} + b6\text{Ethnicity1} + b7\text{Ethnicity2} + b8\text{BeneStatus1} + b9\text{BeneStatus2} + b10\text{BeneStatus3}$
 6. $Y6(\text{Rad Proc}) = a_0U + b1\text{GrpMembership} + b2\text{Age1} + b3\text{Age2} + b4\text{Gender} + b5\text{MarStatus} + b6\text{Ethnicity1} + b7\text{Ethnicity2} + b8\text{BeneStatus1} + b9\text{BeneStatus2} + b10\text{BeneStatus3}$
-